

Raw Sequence Listing Error Summary

ERROR DETECTED SUGGESTED CORRECTION

SERIAL NUMBER: 09/549,827

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 Wrapped Nucleics The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 2 Invalid Line Length The rules require that a line not exceed 72 characters in length. This includes white spaces.
- 3 Misaligned Amino The numbering under each 5th amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
- 4 Non-ASCII The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
- 5 Variable Length Sequence(s) contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6 PatentIn 2.0 A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
- 7 Skipped Sequences Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence:
(OLD RULES) (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
 (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)
 (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
 This sequence is intentionally skipped

Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
- 8 Skipped Sequences Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence.
(NEW RULES) <210> sequence id number
 <400> sequence id number
 000
- 9 Use of n's or Xaa's Use of n's and/or Xaa's have been detected in the Sequence Listing.
(NEW RULES) Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.
 In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 10 Invalid <213> Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or
Response scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
- 11 Use of <220> Sequence(s) 1-4 missing the <220> "Feature" and associated numeric identifiers and responses.
 Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.
 (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12 PatentIn 2.0 Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file,
"bug" resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.

1631

RAW SEQUENCE LISTING

DATE: 07/18/2001

PATENT APPLICATION: US/09/549,827

TIME: 12:44:53

Input Set : A:\A31869A 09549827 wp seq listing.txt

Output Set: N:\CRF3\07182001\I549827.raw

Does Not Comply
Corrected Diskette Needed

pp 1,5

3 <110> APPLICANT: Rzhetsky, Andrey
 4 Kalachikov, Sergey
 5 Krauthammer, Michael
 6 Friedman, Carol
 7 Kra, Pauline
 9 <120> TITLE OF INVENTION: GENE DISCOVERY THROUGH COMPARISONS OF
 10 NETWORKS OF STRUCTURAL AND FUNCTIONAL RELATIONSHIPS AMONG
 11 GENES AND PROTEINS
 14 <130> FILE REFERENCE: A31869-A 070050.104
 C--> 16 <140> CURRENT APPLICATION NUMBER: US/09/549,827
 17 <141> CURRENT FILING DATE: 2000-04-14
 19 <160> NUMBER OF SEQ ID NOS: 22
 21 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 23 <210> SEQ ID NO: 1
 24 <211> LENGTH: 39
 25 <212> TYPE: DNA
 26 <213> ORGANISM: Unknown
 W--> 28 <220> FEATURE:
 W--> 28 <223> OTHER INFORMATION:
 28 <400> SEQUENCE: 1
 29 agcaactaaa caccatcca agcaaacaca cacacaaac 39
 31 <210> SEQ ID NO: 2
 32 <211> LENGTH: 40
 33 <212> TYPE: DNA
 34 <213> ORGANISM: Unknown
 W--> 36 <220> FEATURE:
 W--> 36 <223> OTHER INFORMATION:
 36 <400> SEQUENCE: 2
 37 aagcaactaa acacccatcc aagcaaacac acacacaaac 40
 39 <210> SEQ ID NO: 3
 40 <211> LENGTH: 292
 41 <212> TYPE: DNA
 42 <213> ORGANISM: Unknown
 W--> 44 <220> FEATURE:
 W--> 44 <223> OTHER INFORMATION:
 44 <400> SEQUENCE: 3
 45 aagtacagat ccacggaagg aacgatccaa acaaagacgc aacgacagaa ataacgatcc 60
 46 acataactat ccaaatacat acgcacggaa gtacacacgt aattaaacac ggaagtacat 120
 47 acagatccat ccacggatcc aaataacgaa ttaattacgc atccaaacaa atacggaagt 180
 48 actcaaacac ggaacgaacc atccacggaa ggaacctacat acgtaagcaa ggatccacgg 240
 49 aaggaacgaa gtacctatcc aaacacagac ggaagtaagc aacgacagat cc 292
 51 <210> SEQ ID NO: 4
 52 <211> LENGTH: 10
 53 <212> TYPE: DNA
 54 <213> ORGANISM: Artificial Sequence
 W--> 56 <220> FEATURE:
 W--> 56 <223> OTHER INFORMATION:

see item 11 on Ena Summary Sheet

same env

W

RAW SEQUENCE LISTING

DATE: 07/18/2001

PATENT APPLICATION: US/09/549,827

TIME: 12:44:53

Input Set : A:\A31869A 09549827 wp seq listing.txt

Output Set: N:\CRF3\07182001\I549827.raw

```

56 <400> SEQUENCE: 4
57 atctgtcacg                                     10
59 <210> SEQ ID NO: 5
60 <211> LENGTH: 405
61 <212> TYPE: DNA
62 <213> ORGANISM: Human
64 <400> SEQUENCE: 5
65 catggcttcc tggacaccaa ccctgccatc cgggagcaga cgggtcaagtc catgctgctc 60
66 ctggccccaag agctgaacga ggccaacctc aatgtggagc tgatgaagca ctttgacagg 120
67 ctacaggcca aggatgaaca gggcccccac cgctgcaaca ccacagtctg cctgggcaaa 180
68 atcggtctct acctcagtgc tagcaccaga cacagggtcc ttacctctgc cttcagccga 240
69 gccactaggg acccgtttgc accgtcccggt gttgcgggtg tcctgggctt tgctgccacc 300
70 cacaacctct actcaatgaa cgactgtgcc cagaagatcc tgctgtgct ctgcgggtctc 360
71 actgtagatc ctgagaaatc cgtgcgagac caggccttca aggca 405
73 <210> SEQ ID NO: 6
74 <211> LENGTH: 453
75 <212> TYPE: DNA
76 <213> ORGANISM: Human
78 <220> FEATURE:
79 <221> NAME/KEY: variation
80 <222> LOCATION: (146)...(146)
81 <223> OTHER INFORMATION: A, C, G, or T
83 <400> SEQUENCE: 6
84 ccttcgagtt cggcaatgct ggggccgttg tcctcacgcc cctcttcaag gtgggcaagt 60
85 tcctgagcgc tgaggagtat cagcagaaga tcacccctgt ggtgggtcaag atgtttctcat 120
86 ccactgaccg ggccatgcgc atccgctcc tgcagcagat ggagcagttc atccagtacc 180
87 ttgacgagcc aacagtcaac acccagatct tccccacagt cgtacatggc ttcttgga 240
88 ccaaccctgc catccgggag cagacgggtc agtccatgct gctcctggcc ccaaagctga 300
89 acgaggccaa cctcaatgtg gagctgatga agcactttgc acggctacag gccaaagatg 360
90 aacaggggcc catccgctgc aacaccacag tctgcctggg caaaatcggc tcctacctca 420
91 gtgctagcac cagacacagg gtccttacct ctg 453
93 <210> SEQ ID NO: 7
94 <211> LENGTH: 1727
95 <212> TYPE: DNA
96 <213> ORGANISM: Human
98 <400> SEQUENCE: 7
99 cagccgaagc amgcaaaaat tcttcaggga gctgagcaag agcctggaag cattccctga 60
100 ggayttctgt cggcacaagg tgetgcccc a gctgctgacc gccttcgagt tcggcaatgc 120
101 tggggccggt gtcttcacgc cctcttcaa ggtgggcaag ttctgagcgc ctgaggagta 180
102 tcagcagaag atcatccctg tgggtgtcaa gatgttctca tccactgacc gggccatgcg 240
103 catccgcctc ctgcagcaga tggagcagtt catccagtac cttgacgagc caacagtcaa 300
104 cccccagatc tccccccacg tegtacatgg cttcctggac accaaccctg ccatccggga 360
105 gcagacggtc aagtccatgc tgetcctggc cccaaagctg aacgaggcca acctcaatgt 420
106 ggagctgatg aagcactttg cacggctaca ggccaaggat gaacaggggc ccatccgctg 480
107 caacaccaca gtctgcctgg gcaaaatcgg ctcctacctc agtgctagca ccagacacag 540
108 ggtccttacc tctgccttca gccgagccac tagggaccgg tttgcaccgt cccgggttgc 600
109 ggggtgtcct ggctttgctg ccaccacaaa cctctactca atgaacgact gtgccagaa 660
110 gatcctgcct gtgctctgcg gtctcactgt agatcctgag aaatccgtgc gagaccaggc 720
111 cttcaaggcm wttcgagct tcctgtccaa attggagctc gtgtcggagg acccgacca 780

```

RAW SEQUENCE LISTING

DATE: 07/18/2001

PATENT APPLICATION: US/09/549,827

TIME: 12:44:53

Input Set : A:\A31869A 09549827 wp seq listing.txt

Output Set: N:\CRF3\07182001\I549827.raw

```

112 gctggaggaa gtggagaagg atgtccatgc agcctccagc cctggcatgg gaggagccgc 840
113 agctagctgg gcaggctggg cgtgaccggg gtctcctcac tcacctccaa gctgatccgt 900
114 tcgcacccaa ccactgcccc aacagaaacc aacattcccc aaagaccac gcctgaagga 960
115 gttcctgccc cagccccac ccctgttctt gccacccta caacctcagg ccactgggag 1020
116 acgcaggagg aggacaagga cacagcagag gacagcagca ctgctgacag atgggacgac 1080
117 gaagactggg gcagcctgga gcaggaggcc gagtctgtgc tggcccagca ggacgactgg 1140
118 agcaccgggg gccaaagtga cegtgtagt caggtcagca actccgacca caaatcctcc 1200
119 aaatccccag agtccgactg gagcagctgg gaarctgagg gctcctggga acagggctgg 1260
120 caggagccaa gctcccagga gccacctyct gacggtacac ggctggccag cgagtataac 1320
121 tggggtggcc cagagtccag cgacaagggc gacctcttcg ctacctgtc tgcacgtccc 1380
122 agcaccagc cgaggccaga ctcttggggg gaggacaact gggagggcct cgagactgac 1440
123 agtcgacagg tcaaggctga gctggcccgg aagaagcgcg aggagcggcg gcgggagatg 1500
124 gaggccaaac gcgccgagag gaaggtgcc aaggcccat gaagctggga gcccggaagc 1560
125 tggactgaac cgtgggggtg gcccttcccg gctgcccaga gcccgcccca cagatgtatt 1620
126 tattgtacaa accatgtgag cccggccgcc cagccaggcc atctcacgtg tacataatca 1680
127 gagccacaat aaattctatt tcacaaaaaa aaaaaaaaaa aaaaaaa 1727

```

129 <210> SEQ ID NO: 8

130 <211> LENGTH: 287

131 <212> TYPE: PRT

132 <213> ORGANISM: Human

134 <220> FEATURE:

135 <221> NAME/KEY: VARIANT

136 <222> LOCATION: (4)...(4)

137 <223> OTHER INFORMATION: Any amino acid

139 <221> NAME/KEY: VARIANT

140 <222> LOCATION: (244)...(244)

141 <223> OTHER INFORMATION: Any amino acid

143 <400> SEQUENCE: 8

144 Ser Arg Ser Xaa Gln Lys Phe Phe Gln Glu Leu Ser Lys Ser Leu Asp

145	1		5			10			15
146	Ala	Phe	Pro	Glu	Asp	Phe	Cys	Arg	His
147			20			25			30
148	Thr	Ala	Phe	Glu	Phe	Gly	Asn	Ala	Gly
149			35			40			45
150	Phe	Lys	Val	Gly	Lys	Phe	Leu	Ser	Ala
151			50			55			60
152	Ile	Pro	Val	Val	Val	Lys	Met	Phe	Ser
153	65					70			75
154	Ile	Arg	Leu	Leu	Gln	Gln	Met	Glu	Gln
155						85			90
156	Pro	Thr	Val	Asn	Thr	Gln	Ile	Phe	Pro
157						100			105
158	Asp	Thr	Asn	Pro	Ala	Ile	Arg	Glu	Gln
159						115			120
160	Leu	Ala	Pro	Lys	Leu	Asn	Glu	Ala	Asn
161						130			135
162	His	Phe	Ala	Arg	Leu	Gln	Ala	Lys	Asp
163	145					150			155
164	Asn	Thr	Thr	Val	Cys	Leu	Gly	Lys	Ile

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/549,827

DATE: 07/18/2001

TIME: 12:44:53

Input Set : A:\A31869A 09549827 wp seq listing.txt

Output Set: N:\CRF3\07182001\I549827.raw

165 165 170 175
 166 Thr Arg His Arg Val Leu Thr Ser Ala Phe Ser Arg Ala Thr Arg Asp
 167 180 185 190
 168 Pro Phe Ala Pro Ser Arg Val Ala Gly Val Leu Gly Phe Ala Ala Thr
 169 195 200 205
 170 His Asn Leu Tyr Ser Met Asn Asp Cys Ala Gln Lys Ile Leu Pro Val
 171 210 215 220
 172 Leu Cys Gly Leu Thr Val Asp Pro Glu Lys Ser Val Arg Asp Gln Ala
 173 225 230 235 240
 174 Phe Lys Ala Xaa Arg Ser Phe Leu Ser Lys Leu Glu Ser Val Ser Glu
 175 245 250 255
 176 Asp Pro Thr Gln Leu Glu Glu Val Glu Lys Asp Val His Ala Ala Ser
 177 260 265 270
 178 Ser Pro Gly Met Gly Gly Ala Ala Ala Ser Trp Ala Gly Trp Ala
 179 275 280 285
 182 <210> SEQ ID NO: 9
 183 <211> LENGTH: 223
 184 <212> TYPE: PRT
 185 <213> ORGANISM: Human
 187 <400> SEQUENCE: 9
 188 Val Met Glu Leu Leu Glu Glu Asp Leu Thr Cys Pro Ile Cys Cys Ser
 189 1 5 10 15
 190 Leu Phe Asp Asp Pro Arg Val Leu Pro Cys Ser His Asn Phe Cys Lys
 191 20 25 30
 192 Lys Cys Leu Glu Gly Ile Leu Glu Gly Ser Val Arg Asn Ser Met Trp
 193 35 40 45
 194 Arg Pro Ala Pro Phe Lys Cys Pro Thr Cys Arg Lys Glu Thr Ser Ala
 195 50 55 60
 196 Thr Gly Ile Asn Ser Leu Gln Val Asn Tyr Ser Leu Lys Gly Ile Val
 197 65 70 75 80
 198 Glu Lys Tyr Asn Lys Ile Lys Ile Ser Pro Lys Met Pro Val Cys Lys
 199 85 90 95
 200 Gly His Met Gly Gln Pro Leu Asn Ile Phe Cys Leu Thr Asp Met Gln
 201 100 105 110
 202 Leu Ile Cys Gly Ile Cys Ala Thr Arg Gly Glu His Thr Lys His Val
 203 115 120 125
 204 Phe Cys Ser Ile Glu Asp Ala Tyr Ala Gln Glu Arg Asp Ala Phe Glu
 205 130 135 140
 206 Ser Leu Phe Gln Ser Phe Glu Thr Trp Arg Arg Gly Asp Ala Leu Ser
 207 145 150 155 160
 208 Arg Leu Asp Thr Met Glu Thr Ser Lys Arg Lys Ser Leu Gln Leu Met
 209 165 170 175
 210 Thr Lys Asp Ser Asp Lys Val Lys Glu Phe Phe Glu Lys Leu Gln His
 211 180 185 190
 212 Thr Leu Asp Gln Lys Lys Asn Glu Ile Leu Ser Asp Phe Glu Thr Met
 213 195 200 205
 214 Lys Leu Ala Val Met Gln Ala Tyr Asp Pro Glu Ile Asn Lys Leu
 215 210 215 220
 218 <210> SEQ ID NO: 10

RAW SEQUENCE LISTING

DATE: 07/18/2001

PATENT APPLICATION: US/09/549,827

TIME: 12:44:53

Input Set : A:\A31869A 09549827 wp seq listing.txt

Output Set: N:\CRF3\07182001\I549827.raw

```

219 <211> LENGTH: 218
220 <212> TYPE: PRT
221 <213> ORGANISM: Mouse
223 <400> SEQUENCE: 10
224 Val Leu Glu Met Ile Lys Glu Glu Val Thr Cys Pro Ile Cys Leu Glu
225 1 5 10 15
226 Leu Leu Lys Glu Pro Val Ser Ala Asp Cys Asn His Ser Phe Cys Arg
227 20 25 30
228 Ala Cys Ile Thr Leu Asn Tyr Glu Ser Asn Arg Asn Thr Asp Gly Lys
229 35 40 45
230 Gly Asn Cys Pro Val Cys Arg Val Pro Tyr Pro Phe Gly Asn Leu Arg
231 50 55 60
232 Pro Asn Leu His Val Ala Asn Ile Val Glu Arg Leu Lys Gly Phe Lys
233 65 70 75 80
234 Ser Ile Pro Glu Glu Glu Gln Lys Val Asn Ile Cys Ala Gln His Gly
235 85 90 95
236 Glu Lys Leu Arg Leu Phe Cys Arg Lys Asp Met Met Val Ile Cys Trp
237 100 105 110
238 Leu Cys Glu Arg Ser Gln Glu His Arg Gly His Gln Thr Ala Leu Ile
239 115 120 125
240 Glu Glu Val Asp Gln Glu Tyr Lys Glu Lys Leu Gln Gly Ala Leu Trp
241 130 135 140
242 Lys Leu Met Lys Lys Ala Lys Ile Cys Asp Glu Trp Gln Asp Asp Leu
243 145 150 155 160
244 Gln Leu Gln Arg Val Asp Trp Glu Asn Gln Ile Gln Ile Asn Val Glu
245 165 170 175
246 Asn Val Gln Arg Gln Phe Lys Gly Leu Arg Asp Leu Leu Asp Ser Lys
247 180 185 190
248 Glu Asn Glu Glu Leu Gln Lys Leu Lys Lys Glu Lys Lys Glu Val Met
249 195 200 205
250 Glu Lys Leu Glu Glu Ser Glu Asn Glu Leu
251 210 215
254 <210> SEQ ID NO: 11
255 <211> LENGTH: 124
256 <212> TYPE: PRT
257 <213> ORGANISM: Mouse
259 <400> SEQUENCE: 11
260 Met Glu Pro Val Ala Ser Asn Ile Gln Val Leu Leu Gln Ala Ala Glu
261 1 5 10 15
262 Phe Leu Glu Arg Arg Glu Arg Glu Ala Glu His Gly Tyr Ala Ser Leu
263 20 25 30
264 Cys Pro His His Ser Pro Gly Thr Val Cys Arg Arg Arg Lys Pro Pro
265 35 40 45
266 Leu Gln Ala Pro Gly Ala Leu Asn Ser Gly Arg Ser Val His Asn Glu
267 50 55 60
268 Leu Glu Lys Arg Arg Arg Ala Gln Leu Lys Arg Cys Leu Glu Gln Leu
269 65 70 75 80
270 Arg Gln Gln Met Pro Leu Gly Val Asp Cys Thr Arg Tyr Thr Thr Leu
271 85 90 95

```

Use of n and/or Xaa has been detected in the Sequence Listing.
 Review the Sequence Listing to insure a corresponding
 explanation is presented in the <220> to <223> fields of
 each sequence using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/549,827

DATE: 07/18/2001

TIME: 12:44:54

Input Set : A:\A31869A 09549827 wp seq listing.txt

Output Set: N:\CRF3\07182001\I549827.raw

L:16 M:270 C: Current Application Number differs, Replaced Current Application Number
L:28 M:258 W: Mandatory Feature missing, <220> FEATURE:
L:28 M:258 W: Mandatory Feature missing, <223> OTHER INFORMATION:
L:36 M:258 W: Mandatory Feature missing, <220> FEATURE:
L:36 M:258 W: Mandatory Feature missing, <223> OTHER INFORMATION:
L:44 M:258 W: Mandatory Feature missing, <220> FEATURE:
L:44 M:258 W: Mandatory Feature missing, <223> OTHER INFORMATION:
L:56 M:258 W: Mandatory Feature missing, <220> FEATURE:
L:56 M:258 W: Mandatory Feature missing, <223> OTHER INFORMATION:
L:86 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6
L:144 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8
L:174 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8
L:329 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:14
L:369 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15
L:370 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15
L:373 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15
L:374 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15
L:581 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:21
L:620 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:22